

■研究・実践の課題（テーマ）

食事写真の撮影条件が画像認識に及ぼす影響について

■主任研究者 藤木理代

■共同研究者 福安智哉

■研究・実践の目的、方法、結果、考察や提案等の概要

日常的な健康管理において、栄養摂取量を正確に把握することは重要である。近年、情報処理技術の発展に伴い、スマートフォン等で撮影した食事画像を深層学習により解析し、栄養摂取量を推定する手法が提案されている。しかし、学習用データベースの不足に加え、食事は調理方法や盛り付け方の違いによって外観が大きく変化するため、安定した推定は容易ではない。さらに、食器の形状や撮影条件が推定精度に影響を及ぼすことも知られており、食事画像に基づく栄養摂取量推定には依然として多くの課題が残されている。

日本の食事では、一般に深さのある容器が多く用いられ、また一食当たりの品数も多い傾向にある。先行研究では、食器の深さが単一画像を用いた栄養摂取量推定の精度低下につながる可能性が示唆されているが、これらの知見の多くは諸外国における報告に基づくものであり、日本の食事を対象とした検討は十分ではない。そこで本研究では、日本の食事を対象に、単一画像を用いた栄養摂取量推定において、汎用的な大規模言語モデル間の比較及び盛り付け容器の違いが推定結果に及ぼす影響について検討した。

【方法】

対象食は、作成者の調理技量の影響を排除し、同一条件で繰り返し検討を可能にするため、市販の冷凍弁当である「ベルーナグルメ おまかせ和ごころ御膳」および包装米飯「アイリスオーヤマ パックご飯 180 g」を用いた。栄養価は、サツマイモと豚肉のほくほく煮セットとごはん、エネルギー449kcal、たんぱく質 10.3g、脂質 9.2g、炭水化物 80.3g、塩分量 1.4g、照り焼きハンバーグセットとごはん、エネルギー456kcal、たんぱく質 15.0g、脂質 8.7g、炭水化物 80.1g、塩分量 1.6g であった。購入した食事は、深さのある食器および深さのない食器に盛り付けた。

深さのある容器として、ごはんは深さ 6.0cm、直径 12.0 cm の茶碗を使用した。おかずは、主菜に深さ 2.5cm、幅 16cm の皿、副菜 1 に深さ 2.5cm、幅 7.0cm の皿、副菜 2 に深さ 2.5cm、幅 8cm の皿、副菜 3 に深 1.7cm、幅 5.0cm の皿をそれぞれ使用した。

深さのない容器として、幅 27cm の丸皿を使用した。

撮影時には、料理を 43cm × 30cm の白色ランチマット上に配置し、三脚に固定したスマートフォン（Xperia PRO-I, Sony 製）を用いて、料理に対して 45°の角度から撮影した。撮影設定は、AI による画像補正を行わず、その他の設定は Auto とした。照明条件は、料理直上 100 cm の位置から白色ライトで照射した。

AIによる検証には、ChatGPT 5.4 思考モード、Gemini 3 思考モード、Grok 思考モードの3種類のモデルを用いた。栄養価推定は、各モデルに対し「次の写真から、エネルギー、たんぱく質、脂質、炭水化物、塩分量を推定してください。」という同一の指示文と写真を同時に提示して実施した。思考過程に伴うばらつきを考慮し、同一条件で各検証を3回繰り返した。

得られた結果から、各栄養項目について相対誤差を算出し、モデル間の比較および各モデル内における容器形状の影響を比較した。統計学的有意差の検討にはt検定を用いた。有意水準は5%未満 ($p < 0.05$) とした。

【結果】

各モデルの推定結果は以下の通りである。

「サツマイモと豚肉のほくほく煮セット」において、深さのある容器に盛り付けた写真（以下深皿群）による栄養価推定の平均は、ChatGPT はエネルギー580.0(±62.4)kcal、たんぱく質 17.0(±1.7)g、脂質 15.0(±1.7)g、炭水化物 92.0(±10.4)g、塩分量 2.7(±0.1)g、誤差平均 52%であった。Gemini はエネルギー556.7(±30.6)kcal、たんぱく質 20.7(±2.1)g、脂質 15.3(±0.8)g、炭水化物 84.7(±7.6)g、塩分量 2.5(±0.2)g、誤差平均 55%であった。Grok はエネルギー753.3(±32.1)kcal、たんぱく質 28.0(±3.0)g、脂質 17.0(±5.0)g、炭水化物 117.7(±20.3)g、塩分量 3.8(±1.1)g、誤差平均 108%であった。深さのない容器に盛り付けた写真（以下浅皿群）による栄養価推定の平均は、ChatGPT はエネルギー636.7(±15.3)kcal、たんぱく質 19.7(±1.5)g、脂質 18.0(±0.0)g、炭水化物 94.8(±4.5)g、塩分量 2.6(±0.2)g、誤差平均 67%であった。Gemini はエネルギー603.3(±12.6)kcal、たんぱく質 19.5(±1.3)g、脂質 14.3(±2.5)g、炭水化物 98.5(±9.2)g、塩分量 2.8(±0.3)g、誤差平均 60%であった。Grok はエネルギー743.3(±83.9)kcal、たんぱく質 30.0(±2.6)g、脂質 21.5(±3.1)g、炭水化物 97.3(±9.5)g、塩分量 3.0(±0.5)g、誤差平均 106%であった。

「照り焼きハンバーグセット」において、深皿群の栄養価推定平均は、ChatGPT はエネルギー635.0(±26.0)kcal、たんぱく質 22.3(±1.5)g、脂質 18.3(±0.6)g、炭水化物 93.3(±5.7)g、塩分量 3.2(±0.2)g、誤差平均 63%であった。Gemini はエネルギー610.0(±52.0)kcal、たんぱく質 21.0(±2.6)g、脂質 16.2(±3.3)g、炭水化物 92.7(±6.4)g、塩分量 3.0(±0.2)g、誤差平均 53%であった。Grok はエネルギー734.3(±29.8)kcal、たんぱく質 28.7(±3.2)g、脂質 23.3(±2.9)g、炭水化物 102.0(±5.3)g、塩分量 3.4(±0.5)g、誤差平均 92%であった。浅皿群の栄養価推定平均は、ChatGPT はエネルギー533.3(±23.1)kcal、たんぱく質 25.3(±2.9)g、脂質 14.7(±0.6)g、炭水化物 74.3(±5.8)g、塩分量 2.1(±0.2)g、誤差平均 36%であった。Gemini はエネルギー566.7(±12.6)kcal、たんぱく質 21.5(±1.3)g、脂質 16.0(±0.0)g、炭水化物 83.3(±2.9)g、塩分量 2.6(±0.3)g、誤差平均 44%であった。Grok はエネルギー670.0(±43.6)kcal、たんぱく質 27.3(±1.2)g、脂質 20.3(±0.6)g、炭水化物 102.7(±8.6)g、塩分量 3.9(±0.4)g、誤差平均 87%であった。

各モデル間、または同一モデルにおける深皿群、浅皿群について有意差検定の結果は以

下の通りである。サツマイモと豚肉のほくほく煮セットの深皿群におけるモデル間比較では、ChatGPT と Gemini の間に有意差は認められなかった。Gemini と Grok の間ではエネルギーおよびたんぱく質で有意差を認めた ($p<0.05$)。ChatGPT と Grok の間では、エネルギーおよびたんぱく質に有意差を認めた ($p<0.05$)。浅皿群におけるモデル間比較では、ChatGPT と Gemini の間でエネルギーに有意差を認めた($p<0.05$)。Gemini と Grok の間でエネルギー、たんぱく質、脂質に有意差を認めた ($p<0.05$)。ChatGPT と Grok の間でたんぱく質に有意差を認めた ($p<0.05$)。照り焼きハンバーグセットの深皿群におけるモデル間比較では、ChatGPT 思考モードと Gemini の間に有意差は認められなかった。Gemini と Grok の間で、エネルギー、たんぱく質、脂質に有意差を認めた ($p<0.05$)。ChatGPT と Grok の間で、エネルギー、たんぱく質、脂質に有意差を認めた ($p<0.05$)。照り焼きハンバーグセットの浅皿群におけるモデル間比較、ChatGPT と Gemini の間で脂質に有意差を認めた ($p<0.05$)。Gemini と Grok の間で全項目において有意差を認めた ($p<0.05$)。ChatGPT と Grok の間でエネルギー、脂質、炭水化物、塩分量に有意差を認めた ($p<0.05$)。

深皿群と浅皿群との比較において、サツマイモと豚肉のほくほく煮セットでは、ChatGPT のみ脂質で有意差を認めた($p<0.05$)が、Gemini3 思考モードおよび Grok 思考モードでは有意差を認めなかった。照り焼きハンバーグセットでは、ChatGPT でエネルギー、脂質、炭水化物、塩分量に有意差($p<0.05$)を認めたが、Gemini3 思考モードおよび Grok 思考モードでは有意差を認めなかった。

【考察】

本研究では、汎用的な大規模言語モデルを用いて、単一画像から日本の食事の栄養価推定を行い、容器形状の違いが推定結果に及ぼす影響を検討した。その結果、いずれのモデルにおいても実測値より高く推定される傾向がみられ、特にたんぱく質、脂質、塩分量で過大評価が目立った。一方で、炭水化物量の誤差は比較的小さい傾向を示した。

炭水化物量の誤差が比較的小さかった理由として、各モデルが共通してごはん量をおおむね適切に推定できていたことが考えられる。実際、出力結果では、いずれのモデルもごはん量を 160~180 g 程度、あるいは茶碗 1 杯相当として評価しており、炭水化物の大部分を占める主食のごはんをおおむね適切に評価した結果、炭水化物量全体の推定誤差が比較的抑えられた可能性がある。

一方、本研究では、たんぱく質、脂質、塩分量は過大評価される傾向が認められた。これは、主菜や副菜の調理法、食材の内訳、調味料使用量を単一画像のみから正確に判断することが困難であるためと考えられる。特に、照り、煮汁、ひき肉料理などは、見た目から脂質や調味料使用量が多いと推定されやすく、過大評価につながった可能性がある。

容器形状の影響については、ChatGPT では照り焼きハンバーグセットにおいて、エネルギー、脂質、炭水化物、塩分量に関して推定値の有意な改善が認められた。これは、深さのある容器では料理の重なりや実際の量感が把握しにくく、視覚情報から量を推定する際に誤差が生じやすいのに対し、浅皿にすることでその影響が軽減されたためと考えられる。

しかしながら、サツマイモと豚肉のほくほく煮セットでは、ChatGPTにおいて脂質の推定値が浅皿群で悪化した。これは、浅皿にしたことで料理や煮汁、食材の照りがより強く認識され、脂質として判定された可能性がある。以上のことから、深皿から浅皿への変更は、ハンバーグのように単一のまとまりを形成する料理においては有効である一方、煮物のように複数の食材が広がって配置される料理では有効でない可能性がある。また、今回の実験では、深皿群と浅皿群との比較において、Gemini および Grok では有意差が認められなかった。このことから、容器の深さは一定の条件下では推定結果に好影響を及ぼす可能性があるものの、その影響の程度はモデルごとに異なることが示唆された。

モデル間比較では、Grok は他の2モデルに比べて高値を示す傾向が強く、平均誤差も大きかった。一方、ChatGPT と Gemini は比較的近い値を示す場面が多く、汎用的な AI モデルであっても、推定傾向にはモデル固有の特徴が存在することが示された。

以上のことから、汎用的な AI モデルを用いた単一画像による栄養価推定では、現時点で正確な値を推定することは困難であると考えられる。したがって、日本の食事のように品数が多く、深さのある食器を用いる場面では、単一画像のみで正確な栄養価推定を行うことには限界があり、今後は複数方向からの撮影や、栄養価推定に特化した AI モデルの活用、食器に関する情報などの補足資料を加味することで、より精度の高い推定を行う必要があると考えられる。